

**ПРОГРАММНЫЙ МОДУЛЬ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ
ДЛЯ СИСТЕМ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ****А. А. Мусаев, И. А. Барласов (Санкт-Петербург)**

Применение средств оптимизации управления *сложными динамическими системами* (СДС) предполагает наличие математических моделей, для построения и коррекции которых используются ряды наблюдений, полученные в процессе мониторинга за изменением ее состояния. Однако на практике качество полученного «информационного сырья» оказывается недостаточно для формирования адекватных математических моделей.

В большинстве случаев те или иные алгоритмы прогнозирования и принятия решений базируются на математических методах, применение которых ограничено набором условий. Данные условия обычно относятся к статистическим и динамическим свойствам исходных данных. При невыполнении этих условий соответствующие алгоритмы оптимизации, в лучшем случае, не обеспечивают формирование эффективных решений, а при некоторых обстоятельствах приводят к ложным выводам. В связи с этим применению формализованных методов управления должен предшествовать комплексный анализ свойств информационных потоков и их предварительная обработка. Отсюда непосредственно вытекает рассматриваемая ниже задача разработки *программного модуля* (ПМ), обеспечивающего решение указанных задач.

Кроме того, ПМ обеспечивает решение таких важных для управления задач, как оперативное преобразование потока измерительной информации в совокупность параметров, допускающих наглядную интерпретацию протекающих процессов; выявление скрытой (латентной) информации, описывающей взаимосвязи и взаимовлияния параметров СДС и внешней среды; агрегирование, сжатие и визуализация многомерных взаимосвязанных процессов;

обнаружение аномальных наблюдений и скрытых несоответствий и т. п.

Техническая реализация ПМ может осуществляться в самых разнообразных формах – от программных продуктов, интегрируемых в качестве аналитических надстроек в АСУ, до программных приложений БД. В целом, ПМ предобработки следует рассматривать как функциональный элемент системы поддержки принятия решений, предназначенный для повышения эффективности формируемых управлений.

Концептуальные основы построения подобных модулей и их применения были подробно описаны, например, в [1, 2]. В настоящей публикации рассмотрен конкретный пример реализации ПМ предобработки, ориентированного на задачи анализа данных, полученных в процессе мониторинга состояния ТП первичной переработки нефти.

Функциональная структура программного модуля предобработки (рис. 1)

Результаты мониторинга состояния СДС и ее выходных параметров поступают в БД в виде временной последовательности многомерных данных. Содержащиеся в БД данные могут не в полной мере отвечать требованиям системы обработки данных. В связи с этим возникает необходимость использования модуля предварительной обработки данных, включающая в себя такие функциональности, как обнаружение и обработка пропусков в БД, выявление мультиколлинеарности в исходных данных, сглаживание временных рядов наблюдений и др.

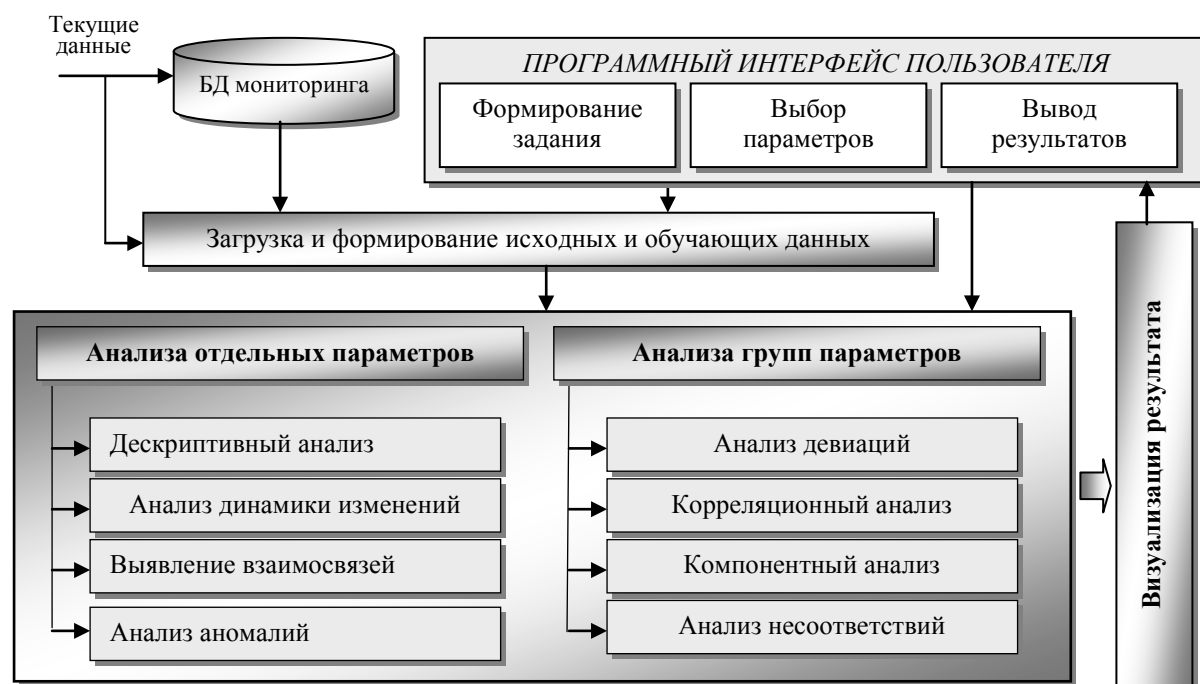


Рис. 1. Функциональная структура модуля предобработки

Выбор загружаемого массива, используемых функциональностей и параметров исследования осуществляется средствами программного интерфейса пользователя (рис. 2). Сюда же, в графические окна интерфейса, выводятся результаты, отвечающие выбранной задаче (функциональности).

Задачи решаются последовательно. При каждом обращении к программе выбирается одна конкретная задача, в соответствии с которой задание передается соответствующей процедуре, отвечающей за реализацию данной функциональности.

Выбор основных функциональностей осуществляется в два этапа. Сначала в блоке выбора «Вид исследований» выбирается либо задача анализа отдельных параметров, либо задача анализа групп параметров. В зависимости от этого выбора активизируется левая или правая половина блока выбора решаемых задач «Основные функциональности», на котором и производится окончательный выбор частной задачи исследований. При этом в случае необходимости, активизируется одно из окон блока выбора параметров анализа «Опции», позволяющего уточнить используемые константы. Всего таких окон три:

окно выбора уровня доверия (используется при решении задач дескриптивного анализа, обнаружения аномальных наблюдений и корреляционных несоответствий);

окно выбора уровня значимости корреляционных связей (используется при решении задачи выявления факторов, имеющих значимые корреляционные связи с исследуемым параметром);

окно выбора порядка полиномиальной аппроксимирующей модели (используется при решении задачи параметрической идентификации динамики изменения исследуемого параметра).



Рис. 2. Внешний вид интерфейса ПМ предобработки данных

Задание на исследование, выбранные параметры и имя файла с массивами исходных данных вводятся в программу кнопкой «Ввод». При этом соответствующие значения высвечиваются для контроля на дисплее монитора. Если введенные данные соответствуют выбору пользователя, то кнопкой «Старт» запускается основная программа исследования, осуществляющая решение поставленной задачи. Соответствующие результаты выводятся в окнах «Результаты анализа» интерфейса. При этом левое окно обеспечивает графический, а правое – текстовый вывод результатов предобработки.

Применение модуля предобработки данных

ПМ может использоваться автономно, на любом рабочем компьютере с современными версиями операционной среды Windows. При этом исходные массивы данных передаются либо в текстовом формате, либо в формате электронных таблиц MS Excel.

Обращение к основной программе ПМ осуществляется через НМІ интерфейс (см. рис. 2). В качестве примеров применения ПМ рассмотрим две функциональности: выявление аномальных наблюдений в БД ретроспективных данных и задачу визуализации многомерного состояния методом компонентного анализа. Оба примера предобработки данных основаны на данных, полученных в процессе мониторинга состояния промышленной технологической установки атмосферной перегонки (ректификации) нефти.

Обнаружение аномальных измерений. При решении задач статистического анализа данных наличие в выборке аномальных измерений может привести к существенной потере в точности формируемых оценок и прогнозов. В связи с этим в ПМ предусмотрена функциональность, позволяющая обнаруживать и устранять аномальные наблюдения.

Для гауссовского распределения погрешностей наблюдений в качестве решающего правила при решении задачи обнаружения аномальных наблюдений можно использовать простейший подход, основанный на попадании наблюдений за пределы доверительного интервала $\Delta_x = (\bar{x} - \delta x, \bar{x} + \delta x)$. При этом вероятность того, что наблюдения лежат в пределах данного интервала обычно выбирается из числового ряда 0,95, 0,99, 0,999, 0,997. Пренебрегая вероятностью соответствующей ошибки, можно рассматривать все наблюдения, не попадающие в Δ_x , как аномальные. Аналогичные результаты дает подход, основанный на проверке статистических гипотез. Одним из распространенных решающих правил является критерий выявления аномальных наблюдений на основе статистики вида [3]

$$B_1' = \frac{x_{(n)} - \bar{x}}{s}, \text{ где } s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Здесь $x_{(n)}$ – «подозрительное» наблюдение, \bar{x} и s – выборочные оценки среднего и ско, построенные по выборке объема n .

В качестве примера на рис. 3 представлен график наблюдений изменения давления в колонне К1 установки АТ6. Звездочками обозначены наблюдения, которые в соответствии с представленным решающим правилом отнесены к категории аномальных для уровня доверия $\alpha = 0.95$.

Компонентный анализ. Для решения задачи визуализации многомерных процессов (порядка $m > 3$) часто используются технологии агрегации данных на основе линейной формы $y_k = \sum_{j=1}^m a_j x_j$, ($k \leq 2 \dots 3$, т. е. допускала бы визуализацию данных), отвечающие условию максимального

сохранения информации, содержащейся в массиве наблюдений X . Один из вариантов решения обеспечивается методом главных компонент [4], в соответствии с которым коэффициенты указанной линейной формы являются собственными векторами A , отвечающими наибольшим собственным числам $\{\lambda_j, j=1, \dots, m\}$ ковариационной матрицы S массива исходных данных X . Для нахождения собственных чисел решается характеристическое уравнение $|S - \lambda I| = 0$, где I – единичная матрица. В качестве примера на рис. 4 приведено представление значений 12-мерного вектора управления технологической установкой АТ6 (установка атмосферной перегонки нефти) в трехмерном пространстве главных компонент.

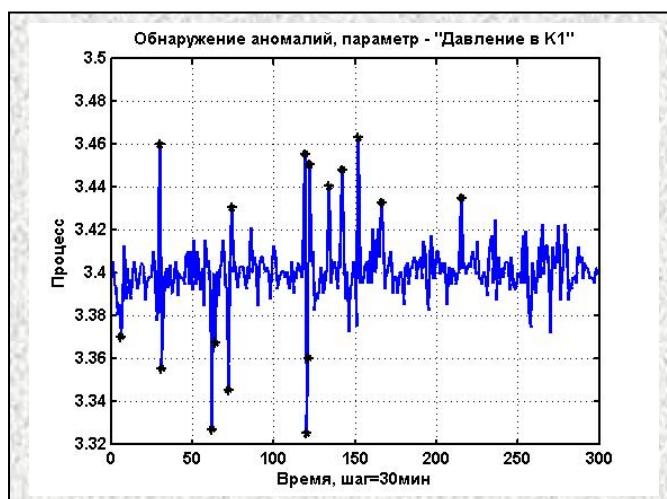


Рис. 3. Результаты обнаружения аномальных наблюдений в ряду наблюдений за параметром «Давление в колонне К1»

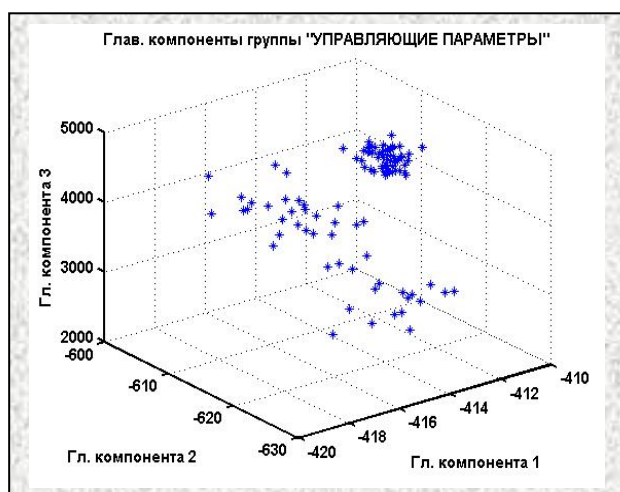


Рис. 4. Представление вектора управления установкой АТ6 в пространстве главных компонент

Выводы

Рассмотренный ПМ предобработки позволяет получить дополнительную информацию о статистических и динамических свойствах контролируемых процессов, взаимосвязях отдельных параметров, получить интегрированное представление об изменении состояния отдельных функциональных групп параметров управления, состояния и выхода. Полученная информация, в свою очередь, дает возможность повысить качество формируемых управлений и получить необходимые сведения для формирования оптимизирующих управлений.

Литература

1. Бахтадзе Н. Н. Виртуальные анализаторы (идентификационный подход)//Автоматика и телемеханика. 2004. № 11. С. 3–24.
2. Мусаев А. А. Виртуальные анализаторы: концепция построения и применения в задачах управления непрерывными технологическими процессами//Автоматизация в промышленности. 2003. № 8. С. 28–33.
3. Введение в теорию порядковых статистик/Под ред. А. Е. Сархана и Б. Г. Гринберга. Пер. с англ. под ред. А. Я. Боярского. М.: Статистика, 1970. 414 с.
4. Прикладная статистика. Классификация и снижение размерности/С.А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин. Под ред. С. А. Айвазяна. М.: Финансы и статистика, 1989. 607 с.

Заметим, что весовые характеристики представлений отдельных групп наблюдений обеспечивают возможность учета их дисперсионных и корреляционных свойств. Это позволяет обнаруживать аномальные векторные наблюдения, содержащие корреляционные несоответствия внутри исследуемой группы. Иными словами, возникает возможность идентифицировать скрытые аномалии (или векторные наблюдения с деформированными корреляционными связями). При этом в качестве критерия аномальности можно использовать вышеописанные правила обнаружения больших отклонений, применяя их в обобщенном пространстве последовательно по каждой главной компоненте.